Comparative Binding Energy (COMBINE) Analysis of Influenza Neuraminidase−Inhibitor Complexes

Ting Wang and Rebecca C. Wade*

European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany

Received September 6, 2000

Neuraminidase is a surface glycoprotein of influenza viruses that cleaves terminal sialic acids from carbohydrates. It is critical for viral release from infected cells and facilitates viral spread in the respiratory tract. The catalytic active site of neuraminidase is highly conserved in all type A and B influenza viruses, making it an excellent target for antiinfluenza drug design. Indeed, neuraminidase inhibitors have recently become available in the clinic for the treatment of influenza. Here, we describe the use of 3D structures of neuraminidase−inhibitor complexes to derive quantitative structure−activity relationships (QSARs) to aid understanding of the mechanism of inhibition and the discovery of new inhibitors. Crystal structures of neuraminidase−inhibitor complexes were used alongside modeled complexes to derive QSAR models by COMparative BINding Energy (COMBINE) analysis (Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. *J. Med. Chem.* **1995**, *38*, 2681−2691). The neuraminidase proteins studied include type A subtypes N2 and N9 (which have ca. 50% sequence identity) and an active site mutant of the N9 subtype. The inhibitors include sialic acid and benzoic acid analogues with diverse frameworks and substitution groups. By considering the contributions of the protein residues and a key water molecule to the electrostatic and van der Waals intermolecular interaction energies, a predictive and robust QSAR model for binding to type A neuraminidase was obtained. In this QSAR model, 12 protein residues and 1 bound water molecule are highlighted as particularly important for inhibitory activity. This QSAR model provides guidelines for structural modification of current inhibitors and the design of novel inhibitors in order to optimize inhibitory activity.

Introduction

The influenza virus has a long history of afflicting humanity and is still the cause of a major respiratory tract disease. Vaccines are frequently ineffective because of the rapid rate of mutation of influenza viral antigens. Two known antiinfluenza drugs, amantadine and its analogue rimantadine, are only effective against the type A influenza virus and are known to result in many side effects and resistance problems. These compounds act by blocking the ion channel of the influenza A virus protein M2,[1] which does not exist in the influenza B virus. Alternative targets for antiviral agents are the two major surface glycoproteins of the virus, hemagglutinin and neuraminidase (NA). Hemagglutinin is responsible for viral binding to cell receptors. NA is important for viral release from the infected cell and viral transport through the mucus in the respiratory tract. The determination of the crystal structures[2,3] of the two proteins has provided opportunities for the design of new and broad-spectrum antiinfluenza drugs. Design targeted at NA[4−6] has been highly successful and two NA inhibitors designed by structure-based approaches were approved for treatment in the clinic in 1999.

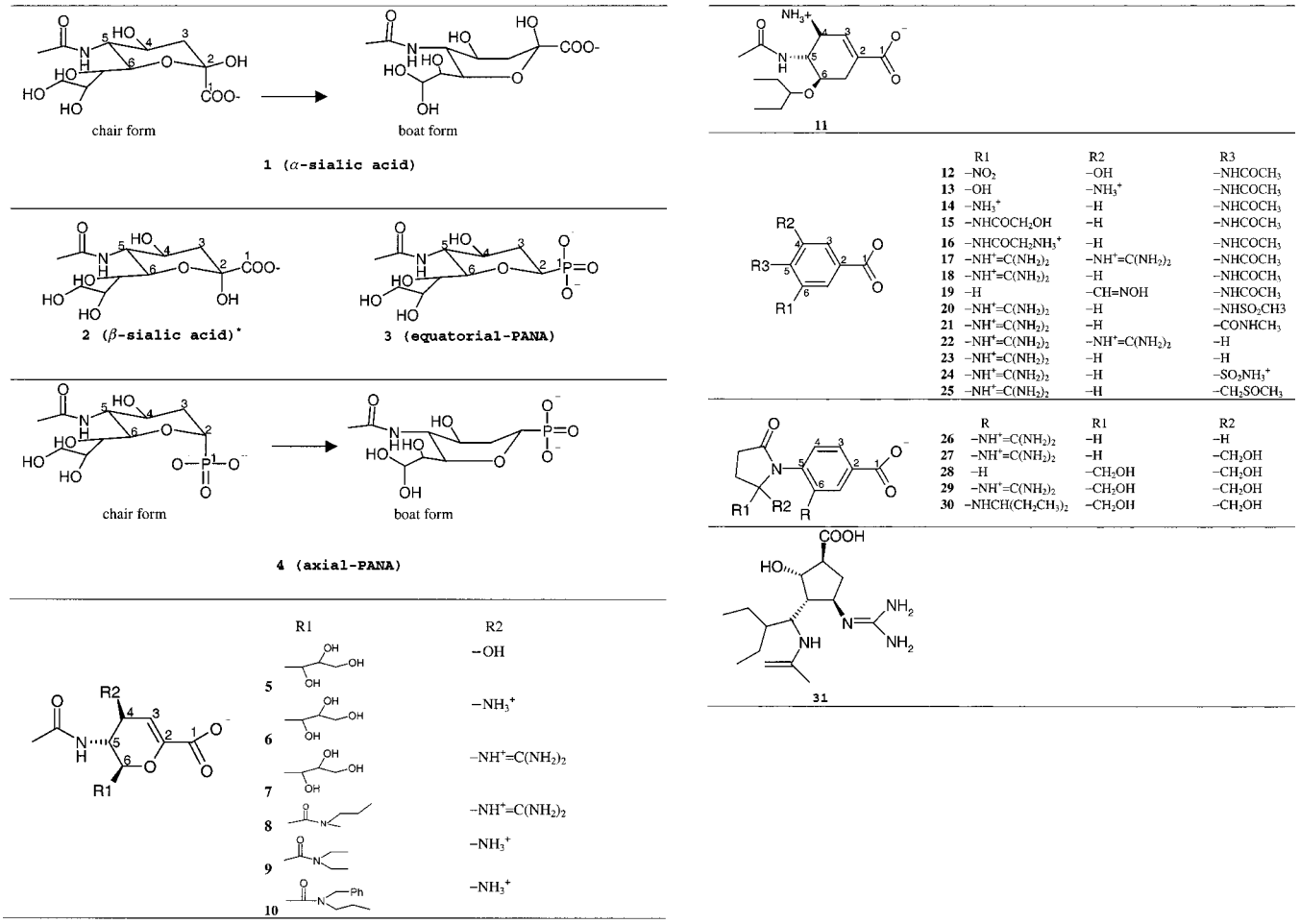
NA[7] is a mushroom-shaped tetrameric protein with a 4-fold symmetry axis. Each monomer fold consists of six antiparallel four-stranded β-sheets. The catalytic active site of NA is located in a concave cavity on the protein surface. Whereas for influenza type A NA, nine subtypes, N1−N9, have been identified, no subtype has been isolated for type B. Despite the considerable diversity of the sequences (e.g. the sequence identity between the N2 A/Tokyo/67 strain and the N9 A/Tern/Australia/G70c/75 strain is 50%), the residues lining the active site of NA are conserved in all wild-type influenza viruses. This makes the active site of NA an excellent target for a broad-spectrum inhibitor. Our study is focused on the N2 and N9 subtypes of type A NA, for which a number of crystal structures of NA−inhibitor complexes have been determined.

Most NA inhibitors currently known are sialic acid analogues[8−12] (see Scheme 1). The substrate of NA, sialic acid (Neu5Ac, **1**), is a weak inhibitor, but its dehydrated transition-state analogue, 2-deoxy-2,3-didehydrosialic acid (Neu5Ac2en, **5**), was found to be 1000-fold more active.[13] After this observation, a series of Neu5Ac2en analogues with higher activities have been reported, one of which, 4-guanidino-Neu5Ac2en (GNA, **7**), has passed through clinical trials and recently came to market as an antiinfluenza nasal spray (Relenza). The ethyl ester prodrug of the achiral inhibitor G39 (**11**) has also passed clinical trials and was recently approved for use as the first orally administered NA inhibitor (Tamiflu). Besides Neu5Ac2en analogues, benzoic acid derivatives[14−17] (see Scheme 1) form another series of potent NA inhibitors that are chemically more accessible. Substitution variants on the benzene ring have led to highly active compounds. Recently a cyclopentane derivative, bcx-

* To whom correspondence should be addressed. Tel: +49 6221 387 553. Fax: +49 6221 387 517. E-mail: wade@embl-heidelberg.de.

10.1021/jm001070j CCC: $20.00
Published on Web 02/10/2001

Scheme 1. Structures of Inhibitors



*Compound **2** (β-sialic acid) does not form a complex with NA.

1812 (**31**), has been shown to have good inhibitory properties,[18] and this compound is undergoing clinical trials. In the following sections, substitution positions in inhibitors are labeled according to the atom number of the pyran ring in sialic acid (**1**) (see Scheme 1): e.g. C4-position for the substitution on the C4 atom. The region of the protein active site near the substitution position is referred to as a pocket with the corresponding label: e.g. C4-pocket for the region of the active site near the C4-position.

Structure-based drug design methods have played a critical role in the discovery of NA inhibitors.[19] The antiinfluenza drug Relenza (**3**), mentioned above, is the direct result of exploiting the crystal structures of NA−inhibitor complexes. The interactions between NA and different probes with various functional groups in the active site[8,20] were analyzed using the GRID program,[21] and the C4-position was revealed to be a highly favorable region for introduction of a basic substituent (e.g. an amino or guanidino group). This group is present in all three drugs that have passed or are undergoing clinical trials.

Quantitative structure−activity relationship (QSAR) studies on NA−inhibitor complexes have also been reported. Taylor and von Itzstein[22] used a molecular dynamics/energy minimization protocol to analyze the structures and predict the relative binding free energies of complexes of a subtype N2 type A NA with a series of C4-substituted Neu5Ac2en inhibitors. Jedrzejas et al.[23] calculated the binding constants (K_i) of NA inhibitors using a continuum solvent electrostatic model and suggested some new benzoic acid compounds with high computed affinities. Recently, Wall et al.[24] reported the application of the linear interaction energy method[25] to computation of binding constants (K_i) for subtype N2 type A NA with Neu5Ac2en analogues. They represented binding free energy differences by a linear combination of electrostatic and van der Waals interaction energies. A knowledge-based PMF scoring function developed by Muegge and Martin[26] was also recently applied to a set of 11 NA−inhibitor complexes in order to score and rank crystallographic and docked complexes.[27] In these computational studies of sets of NA−inhibitor complexes, correlations were achieved between computed binding energies and K_i or IC_{50} values. These were, however, of varying quality and some predictions were incorrect. For example, the compounds predicted to be highly active by Jedrzejas et al.[23] are actually weak inhibitors. In addition, in these studies, the structural diversity of the inhibitors and NA types was limited.

The purpose of the present study is to obtain a high-quality QSAR model that will be a useful guide for the design of potent broad-spectrum NA inhibitors with minimal resistance problems. To achieve this, we apply the COMBINE (COMparative BINding Energy) method[28] and make use of the 3D structures of the protein−ligand

complexes and known IC_{50} values of a set of NA−inhibitor complexes. The set of complexes is chosen to be as large and diverse as possible. It consists of 43 NA−inhibitor complexes containing diverse inhibitors and different subtypes of type A NA. The IC_{50} values for these complexes range over more than 7 orders of magnitude.

In the COMBINE method, the binding free energy ΔG of the receptor−ligand complex (or, in this case, the pIC_{50}) is correlated with a set of selected interaction energy components. Each selected energy component Δu_i contributes to the binding free energy according to its weight w_i:

$$\Delta G = \sum_i w_i \Delta u_i + C \qquad (1)$$

PLS (partial least squares) analysis is applied to obtain the weights w_i. The advantages of this procedure over simply correlating ΔG with the total computed binding energy are that: (1) entropic contributions can be considered explicitly or implicitly in the energy components; (2) the effects of errors in the force-field parametrization or in modeled 3D structures can be, at least in part, filtered out by the PLS analysis; and (3) the resultant model shows which interactions are most important for determining differences in binding affinity and therefore provides hints for the design of molecules with improved binding properties and the estimation of effects due to protein point mutations. The COMBINE method has been successfully applied to inhibitors of phospholipase A_2,[28,29] HIV protease,[30] and glycogen phosphorylase,[31] to substrates of cytochrome P450 1A2,[32] and to study nuclear receptor−DNA binding specificity.[33]

In this study, electrostatic interactions, u_i^{ele}, and van der Waals interactions, u_i^{vdw}, between the inhibitor and each protein residue (and bound water molecules) in energy-minimized structures of NA−inhibitor complexes were selected to estimate the pIC_{50} value:

$$pIC_{50} = \sum_i w_i^{vdw} u_i^{vdw} + \sum_i w_i^{ele} u_i^{ele} + C \qquad (2)$$

The important residues contributing to the activity should exhibit large w_i^{vdw} and/or w_i^{ele} values. Note that a larger (more positive) pIC_{50} value in eq 2 represents stronger binding and corresponds to a more negative binding free energy, ΔG, in eq 1.

A series of NA−inhibitor complexes, including crystal structures and modeled structures, was used to derive COMBINE models. As broad-spectrum activity and minimization of viral resistance are important factors in designing antiinfluenza drugs, subtypes N2 and N9, as well as an active site mutant of subtype N9, were present in the set of NA complexes. This is the first time that COMBINE analysis has been applied to a set of protein−ligand complexes with such large diversity in protein sequence. The NA active site has inherent rigidity, but one of the protein residues (276) is known to undergo a conformational change upon the binding of certain inhibitors and, in addition, the number of tightly bound water molecules is known to vary with the inhibitor. The inhibitors also exhibit structural diversity with different frameworks and substitution groups. The complexity of the NA−inhibitor system is a challenge to the COMBINE method, but also an excellent test. The predictive and robust model obtained not only provides insights useful for the development of new antiinfluenza drugs but also serves as a validation of both the COMBINE method and the docking algorithms used for modeling some of the NA−inhibitor complexes.

Methods and Materials

Abbreviations: Neu5Ac, sialic acid (*N*-acetylneuraminic acid); Neu5Ac2en, 2-deoxy-2,3-didehydro-D-*N*-acetylneuraminic acid; aPANA, 4-acetamido-2,4-dideoxy-D-*glycero*-*β*-D-*galacto*-1-octopyranosylphosphonic acid; ePANA, 4-acetamido-2,4-dideoxy-D-*glycero*-α-D-*galacto*-1-octopyranosylphosphonic acid; BANA, benzoic acid neuraminidase inhibitor; NA, neuraminidase; COMBINE analysis, comparative binding energy analysis; BUW, block unscaled weights; QSAR, quantitative structure−activity relationship; PC, principal component.

Data Set. The crystal structures of 9 N2 and 15 N9 subtype NA−inhibitor complexes were retrieved from the Brookhaven Protein Data Bank. Seven of the 15 N9 complexes have the active site residue mutation Arg292Lys. A further 8 crystal structures of N9 complexes were kindly provided by Dr. Babu of BioCryst Pharmaceuticals, Inc. These contained eight different benzoic acid inhibitors (**18**−**24**) but shared one common set of protein coordinates for a Gly336Asn N9 mutant (n9.pdb). Note that unless otherwise stated, the NA residue numbers given in the text refer to those in the complex with PDB identifier 2qwk.

In addition, one wild-type N9 complex with inhibitor 4AM (**6**) was constructed by superimposing protein atoms from the complex 2qwd (**6**: N9 mutation) and the complex 1nnb (**5**:N9). Also, three complexes of N2 NA with inhibitors GNA (**7**), bcx-140 (**18**), and 4AM (**6**) were constructed by superimposing the complex 1ivf (**5**: N2) and complexes 1nnc (**7**: N9), bc1 (**18**: N9), and 2qwd (**6**: N9 mutant), respectively.

For the inhibitor ST8 (**17**) and the 2-pyrrolidinone-substituted benzoic acid inhibitors (LMA, LMB, LMC, QWM, QWL, **26**−**30**), no crystal structures of complexes with NA were available. The program AUTODOCK3.0[34] was therefore used to flexibly dock these inhibitors into NA in seven complexes. The docking protocol is described below.

In total, 43 complexes containing 29 different inhibitors and two subtypes of NA were prepared for further modeling.

The inhibitory activity data were mainly taken from four references (refs 13−15, 35). IC_{50} values are available for most inhibitors except for the complexes of N2 NA and GNA (**7**) and 4AM (**6**), for which IC_{50} values were calculated from K_i values using the Cheng−Prusoff equation.[31] To normalize experimental data from different literature sources, we took the values in ref 35 as the standard. The IC_{50} values from ref 15 were multiplied by 2 since the IC_{50} value of Neu5Ac2en (**5**) binding to N9 NA in ref 15 is one-half that in ref 35.

The coordinate (PDB) identifiers of the 43 complexes, the protein subtypes, inhibitor names and charges, mutated residues, and pIC_{50} values along with bibliographic sources are listed in Table 1.

Molecular Mechanics Modeling. 1. Preparation of Structures of Complexes. Besides proteins and inhibitors, most crystal structures contained some oligosaccharide groups, two calcium ions (one near to the active site), and a number of ordered water sites. In our modeling studies, the oligosaccharide groups and the calcium ion far from the active site were assumed to be unimportant and were removed. Thus, for each complex, 388 amino acid residues, 1 calcium ion, the inhibitor, and ordered water sites were retained for study.

The cysteines in NA were defined to participate in 9 disulfide bridges. The WHATIF program[37,38] was used to determine the protonation states of histidines and add polar hydrogen atoms to the protein and water by optimization of the hydrogen bond network. All histidines were assigned as singly protonated. As a result, for the N9 subtype, 3 (150, 184,

Table 1. Influenza NA−Inhibitor Complexes

no.	inhibitor:NA[a]	code[b] of complex	mutation	inhibitor name	inhibitor charge (e)	complex source	$pIC_{50}^{exp\ c}$	pIC_{50}^{pred}
1	**5**:N9 mutant	2qwc	Arg292Lys	Neu5Ac2en	−1	PDB	3.39[30]	2.67
2	**6**:N9 mutant	2qwd	Arg292Lys	4AM	0	PDB	4.0[30]	4.58
3	**7**:N9 mutant	2qwe	Arg292Lys	GNA	0	PDB	6.96[30]	5.02
4	**8**:N9 mutant	2qwf	Arg292Lys	G20	0	PDB	5.28[30]	7.07
5	**9**:N9 mutant	2qwg	Arg292Lys	G28	0	PDB	3.64[30]	4.97
6	**11**:N9 mutant	2qwh	Arg292Lys	G39	0	PDB	4.89[30]	4.77
7	**8**:N9	2qwi		G20	0	PDB	7.70[30]	7.26
8	**9**:N9	2qwj		G28	0	PDB	6.64[30]	6.74
9	**11**:N9	2qwk		G39	0	PDB	8.70[30]	8.06
10	**5**:N9	1nnb		Neu5Ac2en	−1	PDB	4.70[30]	3.94
11	**7**:N9	1nnc		GNA	0	PDB	8.70[30]	8.45
12	**10**:N9	1bji		G21	0	PDB	8.70[10]	7.42
13	**3**:N9	1iny	Ser370Leu	ePANA	−2	PDB	3.16[11]	4.27
14	**12**:N2	1ivd		ST1	−1	PDB	3.12[13]	1.65
15	**13**:N2	1ivc		ST2	0	PDB	1.7[13]	1.48
16	**14**:N2	1ive		ST3	0	PDB	1.4[13]	1.44
17	**15**:N2	1ing		ST5	−1	PDB	2.40[13]	4.15
18	**16**:N2	1inh		ST6	0	PDB	2.30[13]	2.45
19	**5**:N2	1ivf		Neu5Ac2en	−1	PDB	4.82[13]	3.90
20	**3**:N2	1inx		ePANA	−1	PDB	4.7[11]	4.32
21	**18**:N9	bc1	Gly336Asn	bcx-140	0	Dr. Babu	5.30[14]	4.24
22	**19**:N9	bc2	Gly336Asn	bcx-384	−1	Dr. Babu	1.96[14]	3.31
23	**20**:N9	bc3	Gly336Asn	bcx-167	0	Dr. Babu	3.70[14]	4.56
24	**21**:N9	bc4	Gly336Asn	bcx-352	0	Dr. Babu	5.00[14]	4.18
25	**22**:N9	bc5	Gly336Asn	bcx-141	+1	Dr. Babu	2.52[14]	3.46
26	**23**:N9	bc6	Gly336Asn	bcx-448	0	Dr. Babu	3.00[14]	3.18
27	**24**:N9	bc7	Gly336Asn	bcx-1023	+1	Dr. Babu	4.74[14]	3.30
28	**25**:N9	bc8	Gly336Asn	bcx-869	0	Dr. Babu	3.55[14]	4.75
29	**26**:N9	lma	Gly336Asn	LMA	−1	AUTODOCK	3.60[12]	4.59
30	**27**:N9	lmb	Gly336Asn	LMB	−1	AUTODOCK	4.70[12]	4.30
31	**28**:N9	lmc	Gly336Asn	LMC	0	AUTODOCK	3.12[12]	4.07
32	**29**:N9	qwm	Gly336Asn	QWM	0	AUTODOCK	5.30[12]	5.55
33	**30**:N9	qwl		QWL	−1	AUTODOCK	7.32[12]	6.89
34	**17**:N9	bc9	Gly336Asn	ST8	+1	AUTODOCK	4.15[15]	5.01
35	**17**:N2	dk3		ST8	+1	AUTODOCK	4.15[15]	4.27
36	**6**:N9	wdd		4AM	0	superposition	6.50[10]	5.84
37	**7**:N2	dk1		GNA	0	superposition	8.81[8]	7.45
38	**18**:N2	dk2		bcx-140	0	superposition	5.00[13]	3.63
39	**6**:N2	dam		4AM	0	superposition	6.41[8]	6.68
40	**1**:N9 mutant	2qwb	Arg292Lys	α-Neu5Ac	−1	PDB	0.70[41]	2.63[d]
41	**1**:N9	1mwe		α-Neu5Ac	−1	PDB	1.70[41]	6.17[d]
42	**1**:N2	2bat	Asp338Asn	α-Neu5Ac	−1	PDB	2.7[13]	6.94[d]
43	**4**:N2	1inw		aPANA	−1	PDB	2.7[11]	4.88[d]
44	**31**:N9	rw9		bcx-1812	0	AUTODOCK	*e*[18]	8.36[d]
45	**31**:N2	rw2		bcx-1812	0	AUTODOCK	*e*[18]	7.13[d]

[a] Inhibitors are shown in Scheme 1. The N9 subtype is referred to as "N9 mutant" if there is a mutation in the active site. [b] PDB identifier if with 4 letters. [c] Experimental data were taken from the references noted. [d] External predicted activities for the N2 + N9 model derived with 39 complexes, whereas pIC_{50}^{pred} values for the first 39 complexes are from cross-validation (see also Figure 2). [e] The pIC_{50} against 15 different strains of type A NA is in the range of 8.85−10.

274) out of 7 histidines were protonated on Hϵ and the other 4 (98, 144, 233, 312) were protonated on Hδ. For the N2 subtype, 7 (144, 150, 155, 168, 184, 264, 274) out of 8 histidines were protonated on Hϵ and 1 (191) on Hδ. The carboxylic acid moieties of the inhibitors were treated as ionized, as were the amino acid side chains of arginine, lysine, aspartate, and glutamate residues. The amino ($-NH_2$) and guanidino groups on the inhibitors were protonated (assuming physiological conditions).

Due to the poorer resolution of some crystals, some structures (1nnb, 1iny, 1inh, 1ivc, 1ive, and 1ing) lack some or all bound water sites. However literature evidence indicates that some water molecules located in the active site are important for NA−inhibitor binding. Therefore, if these water molecules were missing from the structure of a complex, we added them. Complexes with the same or similar proteins and inhibitors should have the same or similar coordinates for bound water molecules; thus structures 2qwc, 1mwe and 1ivd were chosen as templates for adding water molecules by superimposing the proteins. The four bound water molecules in structure 2qwc (41W, 288E, 121G and 1T), which bridge between the Neu5Ac2en inhibitor and the protein, were added to structure 1nnb, and the missing first residue (arginine) in 1nnb was also added. One water molecule (327H) of structure 1mwe was added to structure 1iny. Two water molecules (564H and 486H) of structure 1ivd were added to structures 1inh, 1ivc and 1ive, and one water molecule (564H) of structure 1ivd was added to structure 1ing.

2. Parametrization of Complexes. Parametrization was performed in the xLEaP module of the AMBER5.0 program[39] and InsightII package.[40] For the protein atoms and water molecules, the all-atom AMBER 1995 force-field parameters[41] were used. Some of the parameters required for the inhibitors were not present in this force field and had to be derived. The inhibitors were assigned atomic partial charges according to their AMBER potential types in the InsightII package. This required modification of the AMBER potential template in the InsightII package in order to introduce some new atom types into the force field. In addition, the nitro ($-NO_2$), sulfone ($-SO_2-$) and sulfoxide ($-SO-$) groups had to be assigned atomic partial charges manually from Mulliken MOPAC (MNDO) charges computed within the InsightII package as no existing InsightII potential types were appropriate for these groups. These groups are present in inhibitors bcx-167, bcx-1023, bcx-869 and ST1. Two new atom types, S1 and S2, were defined respectively for the sulfur atoms in sulfone and sulfoxide groups in the AMBER force field. Other parameters for the inhibitors were taken from the CHARMM22 force

field,[42] when not present in the AMBER 1995 force field. The calcium ion was assigned a +2e charge, a 1.74 Å van der Waals radius, and a 0.0465 kcal/mol epsilon value.

3. Energy Minimization of Complexes. Energy minimization of each complex was carried out using the Sander module in AMBER5.0 and consisted of two stages. In the first stage of 200 steps, only hydrogen atoms and water molecules were allowed to move. In the second stage, the protein nonhydrogen atoms were restrained to their crystallographic positions by a harmonic potential with a force constant of 32 kcal/(mol·Å^2) while the hydrogen atoms, the inhibitor and water molecules were unrestrained. The convergence criterion was that the root-mean-square value of the Cartesian elements of the energy gradient was less than 10^{-2} kcal/(mol·Å). A nonbonded cutoff of 10.0 Å and a distance-dependent dielectric constant ($\epsilon = r_{ij}$) were used throughout. In each stage, the first 100 steps were performed with the steepest descent algorithm and the rest of the steps were performed with the conjugate gradient method.

After energy minimization, all bound water molecules, except one involved in interactions between the inhibitors and the protein, were discarded. The retained water molecule is located between the C5- and C6-pockets (e.g. the water 121GH in structure 2qwc). It accepts a hydrogen bond from the C5-position amide hydrogen of the inhibitor and donates hydrogen bonds to Glu277 and Glu227.

4. Docking Protocol for Noncrystallographic Complexes. The same parametrization and minimization protocol was applied to the three complexes derived by superimposition and the seven complexes with ligands docked by AUTODOCK3.0. The docking protocol was as follows.

First, the starting conformation of each inhibitor was built and minimized in the SYBYL software package.[43] Then the conformation was manually adjusted to be as similar as possible to the crystal conformation shown in refs 13, 16. The inhibitors were assigned AMBER atomic partial charges as described in the parametrization section. Nonpolar hydrogen atoms of the inhibitors were removed and their charges were added to those of their bonded carbon atoms. The protein structures also contained only polar hydrogen atoms and were assigned charges with the program q.kollu in AUTODOCK3.0. The Lamarckian genetic algorithm in AUTODOCK3.0 was applied to search the conformational and orientational space of the inhibitors while keeping the protein structures rigid. The default parameters were used. 50 docking runs were performed for each complex. The orientation with the lowest docked energy in the top-ranked cluster was chosen to make a complex for COMBINE analysis.

For each inhibitor, an appropriate protein structure needed to be chosen. This is the consequence of using a rigid protein structure. In this work, the N2 protein of structure 1ivf was used to dock inhibitor ST8 (**17**) and create a N2 complex (dk3). The n9.pdb protein structure was used to dock inhibitors ST8 (**17**), LMA (**26**), LMB (**27**), LMC (**28**) and QWM (**29**) and create five N9 complexes (bc9, lma, lmb, lmc and qwm). The inhibitor QWL (**30**) has a big hydrophobic group at the C6-position and a conformational change is known to take place in residue Glu276 when the C6-position is substituted by a hydrophobic group. Therefore, the N9 protein of structure 2qwk, in which the inhibitor G39 (**11**) has the same hydrophobic group as QWL (**30**), was used to dock QWL (**30**) and create an N9 complex (qml).

After docking, the original bound water sites of each protein structure were added to the docked complex and the water molecules causing a steric clash with the inhibitor were removed. It is worth noting that one of the $-CH_2OH$ groups in inhibitors LMC (**28**), QWM (**29**) and QWL (**30**) expels the important water molecule which was kept for building COMBINE models.

Chemometric Analysis. The COMBINE program (provided by A. R. Ortiz) was used to decompose the interaction energy between the inhibitor and the protein in each minimized complex. That is, it was used to calculate the Lennard−Jones and electrostatic interactions between the inhibitor and each protein residue, the water molecule, and the calcium ion. For complexes lmc, qwm, and qwl containing no water molecule, zero values were set for the energy variables of the water.

These energy descriptors were input into the GOLPE4.5 program[44] for PLS analysis. The variables were divided into two blocks: a van der Waals block and an electrostatic block. Because the variables of the van der Waals block covered a much smaller range than those of the electrostatic block, a block unscaled weights (BUW) pretreatment was used to give the same importance to each block and ensure that the effects of the van der Waals block are not ignored in the QSAR models derived. After the BUW pretreatment, an initial PLS model was built using all the variables and the number of latent variables that produced the most predictive models, as judged by Q^2 values in leave-one-out cross-validation, was chosen as the optimal dimensionality. Then a D-optimal preselection was applied to remove the noise variables by considering the optimal dimensionality in partial weights space. As a result, about 50% of the variables were retained to undergo a further selection by performing fractional factorial design (FFD). The variables unimportant for activity were discarded and the remaining variables were used to build the final PLS model.

To study the robustness of the above procedure, external cross-validation was performed by randomly taking 10 complexes out of the original data set as a prediction set and using the rest as a training set. The training set was used to build a model following the above procedure and to predict the activities in the prediction set. 15 such test models were generated. The complexes in the prediction sets were determined as follows. In the complete data set, pIC_{50} values vary from 1.4 to 8.8, so the complexes were classified into eight activity ranges with an increment of 1.0 from 1.0 to 9.0. One randomly chosen complex per range was assigned to the prediction set, but two complexes were chosen from ranges [3.0−4.0] and [4.0−5.0] because these two ranges contained the majority of the complexes. As a result, a prediction set included 10 randomly chosen complexes with an activity distribution similar to that of the full data set.

Results and Discussions

Model for N2 and N9 NA Subtypes. Although both N2 and N9 NA subtypes contain 388 amino acid residues and conserved active site residues, they differ in most positions in their sequence alignment. The alignment of subtypes N2 and N9 revealed 10 gap residues: 5 in N2 and 5 in N9. The interactions of these gap residues were not considered for the COMBINE analysis. Thus 383 residues, 1 calcium ion, and 1 water were kept in both N2 and N9 complexes for the COMBINE interaction energy calculation, and therefore, 770 *X*-variables were input to the GOLPE4.5 program.

First, all the N2 and N9 complexes except 2qwb, 1mwe, 2bat, and 1inw, i.e. 39 complexes, were used to build a PLS model. The maximum dimensionality for subsequent *X*-variable selection was determined to be five latent variables. After variable selection, about 330 variables were selected to build the final PLS model. The statistical parameters of the model and the average SDEP for the 15 external test sets are listed in Table 2. The dependence of R^2 and Q^2 values on the number of latent variables in the model is shown in Figure 1. The R^2 and Q^2 curves develop slowly beyond three latent variables. The average SDEP value for external validation is, as expected, larger than that for internal validation, but sufficient to demonstrate the robustness of the model. The predicted pIC_{50} values are plotted against experimental pIC_{50} values for the model with three latent variables in Figure 2. Comparison of

Table 2. Predictive Performances of COMBINE Models[a]

data set	no. of complexes	no. of LV	R^2	SDEC	Q^2	SDEP	SDEP^{ext} ± SD
N2 + N9	39	3	0.88	0.69	0.77	0.96	1.22 ± 0.20
		4	0.89	0.64	0.78	0.94	1.20 ± 0.22
		5	0.92	0.57	0.78	0.94	1.19 ± 0.21
N9	28	3	0.91	0.56	0.76	0.93	
		4	0.94	0.45	0.84	0.76	
		5	0.95	0.42	0.85	0.72	

[a] Abbreviations: LV, latent variable; R^2, correlation coefficient; SDEC, standard deviation of errors of correlation; Q^2, predictive correlation coefficient; SDEP, standard deviation of errors of prediction; SDEP^{ext}, average SDEP of 15 external validation tests; SD, standard deviation.

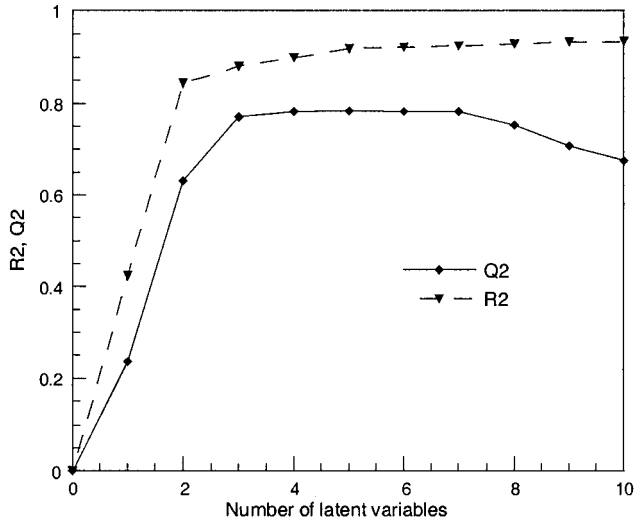


Figure 1. Dependence of the R^2 and Q^2 values on the number of latent variables for the N9 + N2 COMBINE model for 39 complexes.

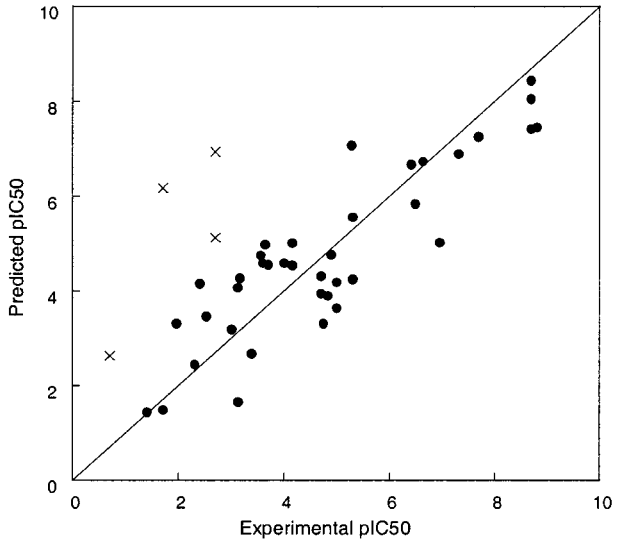


Figure 2. Plot of experimental pIC_{50} values versus predicted pIC_{50} values for the N2 + N9 model derived for 39 complexes: ●, predicted values for the 39 complexes from leave-one-out cross-validation at three latent variables; ×, external predicted values for sialic acids (**1**) in structures 2qwb, 1mwe, and 2bat and for aPANA (**4**) in structure 1inw.

measures of predictive ability shows that, overall, the COMBINE model produces more accurate predictions than obtained in previous SAR studies on NA inhibitors. This is achieved for the largest and most diverse set of

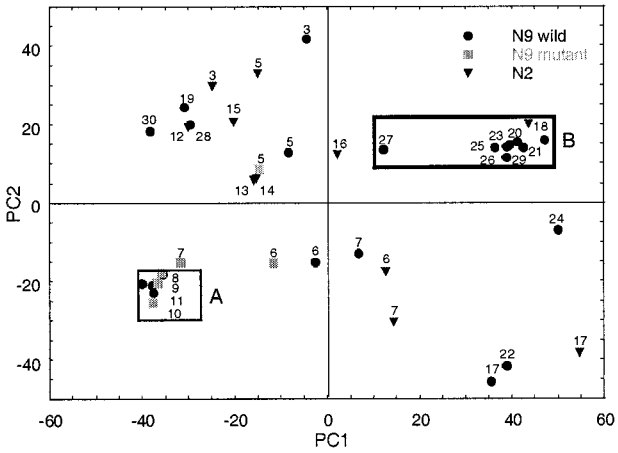


Figure 3. Score plot of the first (PC1) and second (PC2) principal components for the 39 N9 + N2 complexes. Each complex is labeled by the inhibitor number. Neu5Ac2en inhibitors with both C4-position amino groups and C6-position hydrophobic groups cluster in box A; benzoic acid inhibitors with C6-position guanidino groups and neutral groups in other ring positions cluster in box B.

NA−inhibitor complexes studied computationally to date.

To investigate the distribution of the complexes in the space defined by their interaction energies, a principal component analysis was performed for the 39 complexes using the BUW pretreated variables. The score plot of the first two principal components (PC1 and PC2) is shown in Figure 3. PC1 distinguishes the benzoic acid inhibitors with guanidino groups from the others, while PC2 distinguishes inhibitors with positively charged groups at the C4-position from others. The Neu5Ac2en inhibitors with both amino groups at the C4-position and hydrophobic groups at the C6-position closely cluster together. So too do the benzoic acid inhibitors with guanidino groups at the C6-position and neutral groups at other ring positions. No obvious distinction between N2 and N9 complexes can be seen.

To investigate the contributions of the *X*-variables to the activity (pIC_{50}), the partial weights in the first four latent variables (LV1−4) are shown in Figure 4. It can be seen in Figure 4a that the first two latent variables are predominantly defined by van der Waals interactions from Arg118, Glu119, Asp151, Arg152, Trp178, Ile222, Arg224, Ala246, Arg292, Asn294, Arg371, and the water molecule and electrostatic interactions from Asp151, Arg156, Trp178, Arg292, Arg371, and the water molecule. All these residues are located in either the first or the second shell of the active site. The third and fourth latent variables are also mainly defined by these residues, but in addition, residues Ser179, Lys350, Glu425, and the calcium ion contribute. The most important variable for LV4, the Arg152 van der Waals interaction, is also important for LV3 and makes a much larger contribution to LV4 than all other interactions. Therefore, the addition of the fourth latent variable to the model results in little improvement in the predictive ability of the model.

Although the partial weights plots can highlight the variables important for activity, their actual contributions to the PLS equation are not clear. According to eq 2, pIC_{50} values are mainly determined by large PLS coefficients w_i and large interaction energies u_i. In the

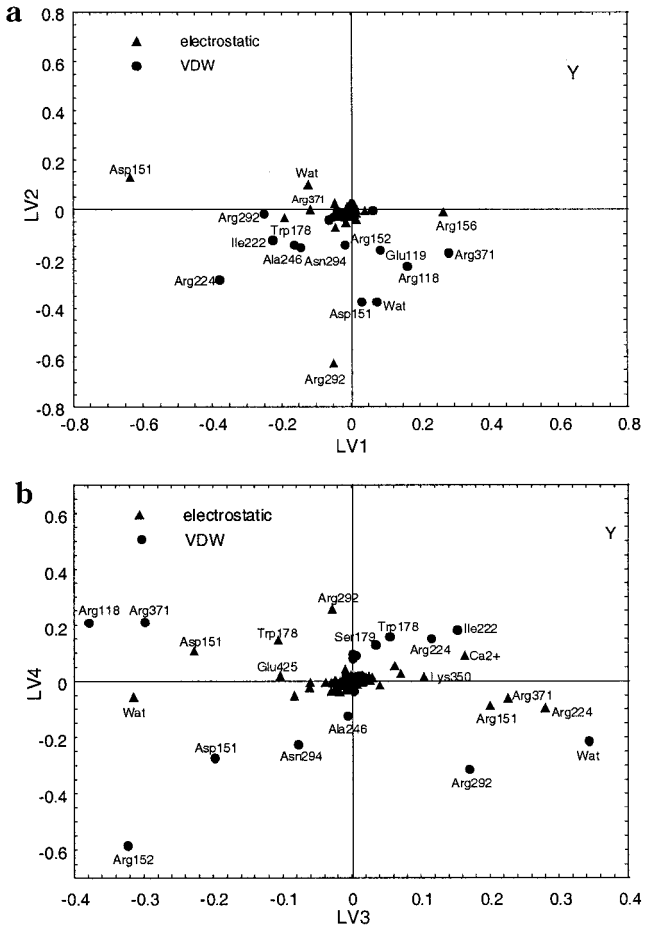


Figure 4. Partial weights plot of the *X*-variable energy terms and the *Y*-variable activity for the PLS latent variables: (a) first (LV1) and second (LV2); (b) third (LV3) and fourth (LV4).

model at three latent variables, the maximum real PLS coefficients of van der Waals and electrostatic interactions are 0.58 and 0.088, respectively. We used a threshold of 0.1 on the PLS coefficients to extract important van der Waals variables with large PLS coefficients and a threshold of 0.01 on the PLS coefficients for important electrostatic variables. The van der Waals variables above the threshold are from residues Arg118, Asp151, Arg152, Trp178, Ile222, Arg224, Ala246, Arg292, Asn294, and the water molecule. The electrostatic variables above the threshold are from residues Asp151, Arg156, Trp178, Arg292, Arg371, Tyr406, and the water molecule. These residues are also the residues that have relatively larger interaction energies with inhibitors. This can be seen in Figure 5, which shows the electrostatic (Figure 5a) and van der Waals (Figure 5b) interaction energies between the inhibitor G39 (**11**) and N9 residues in the structure 2qwk. Figure 5c,d shows the product of the interaction energies and the PLS real coefficients (Figure 5c for the electrostatic term and Figure 5d for the van der Waals term).

Therefore, the activity differences of the inhibitors will be predominantly explained by their interactions with the 12 protein residues listed above and the water molecule. Some of these interactions, e.g. with Asp 151 and Arg 292, are particularly important.

Arg118, Arg292, and Arg371 are located in the C1-pocket and form the triarginyl cluster that anchors the carboxylate groups of the inhibitors. Moreover, Arg371 forms salt links to the carboxylate groups of most inhibitors; this is why Arg371 has a positive van der Waals interaction energy with the inhibitor, as shown in Figure 5b. Asp151 is located in the C4-pocket and makes direct favorable electrostatic interactions with positively charged groups at the C4-position. Arg156 is near Asp151, but in the second shell of the active site. It increases binding affinity by its favorable electrostatic interaction with Asp151 even though its interaction with the inhibitors is electrostatically unfavorable. Therefore, Arg156 has a positive PLS coefficient while Asp151 has a negative PLS coefficient. Arg152 is in the C5-pocket and donates a hydrogen bond to the amide oxygen at the C5-position of the inhibitor. Trp178 and Ile222 as well as the side chain of Arg224 form a small hydrophobic region at the C5-pocket. For inhibitors with a large hydrophobic group at the C6-position, Glu276 rotates and forms a salt link to Arg224, thus creating a necessary hydrophobic pocket together with Ala246 to accommodate the inhibitor. Asn294 also makes contributions to this hydrophobic interaction. Tyr406 lies under the backbone ring at the C2-pocket and might make aromatic interactions with the benzene rings of benzoic acid inhibitors by its phenyl ring and also electrostatic interactions with the ring oxygen of sialic acid inhibitors by its hydroxyl group. In addition, the remarkable significance of Arg292 implies that its mutation will result in a large effect on activity: All the inhibitors exhibited reduced binding to the Arg292Lys mutant NA. This may be explained by noting that, although Lys292 conserves the positive charge of Arg292, Lys292 fails to form a hydrogen bond to the carboxylate group of the inhibitor in the structures of the complexes of the mutant. A stereoview of all the important residues is shown in Figure 6.

It is interesting to note that Glu119 and Glu227 are not significant in the above model although they are in the active site and show strong interactions with the inhibitors (see Figure 5). It appears that the variance in activity in this data set can be described without taking these two residues into account. However, this does not mean that they are not important for activity; indeed, they might contribute to the constant term in the PLS equation.

Blick et al.[45] analyzed the effects of the Glu119Gly mutation on the inhibitory activities of Neu5Ac2en analogues, and the results showed that the inhibitory activity of 4-guanidino-Neu5Ac2en (**7**) was reduced by 500-fold but that of 4-amino-Neu5Ac2en (**6**) remained unchanged. This illustrates that Glu119 is not necessary for the high affinity of an inhibitor to NA although it might be important for some large positively charged substituents, e.g. guanidino at the C4-position. Glu227 is located on the floor of the active site under the C5-position and could make long-range charge−charge interactions with the positively charged groups at the C4-position.

Model for N9 NA Subtype. To investigate whether any important differences exist between the N9 and N2 complexes, 28 N9 complexes were extracted from the data set and used to build a PLS model. The same protocol as applied to the whole data set was used except that two cycles of FFD were performed to obtain the

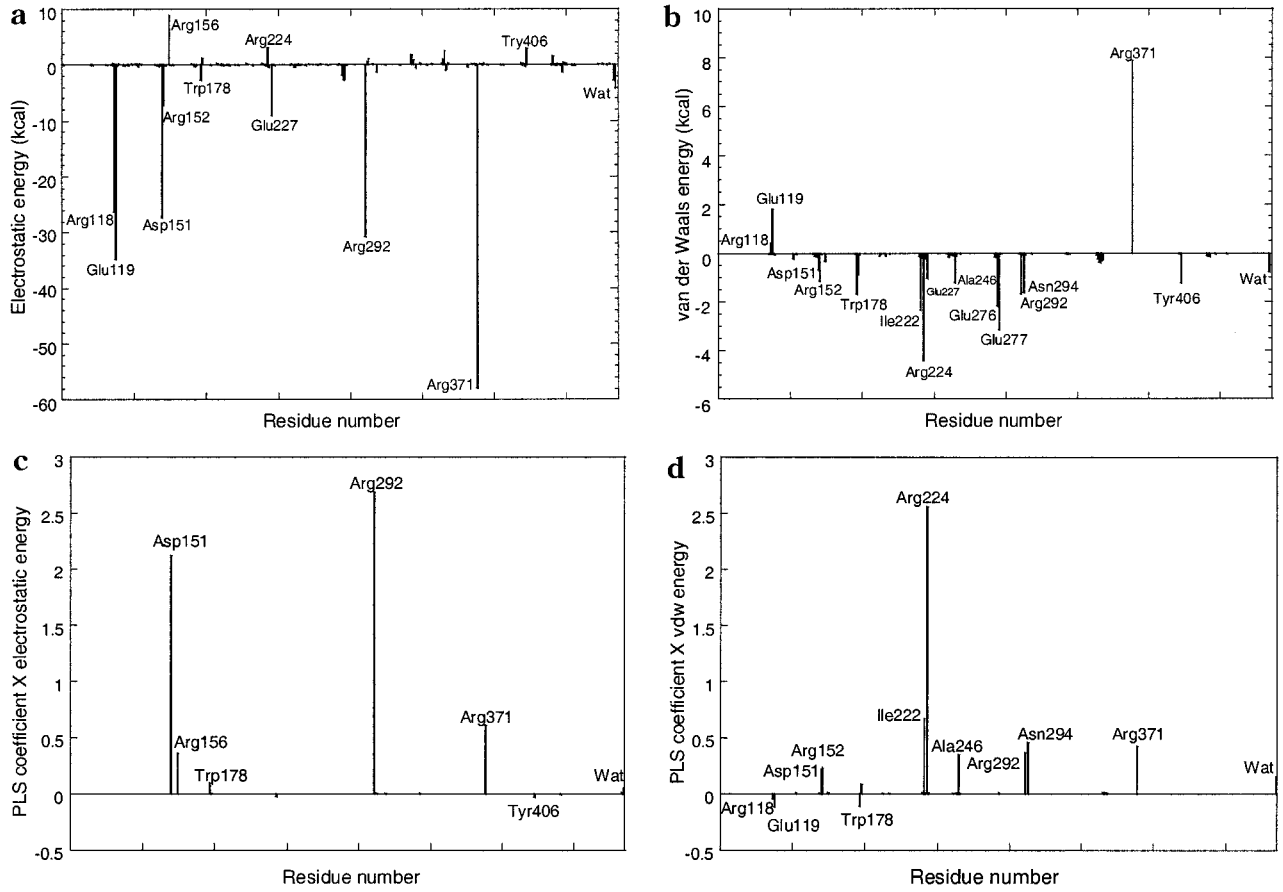


Figure 5. Electrostatic (a) and van der Waals/Lennard−Jones (b) interaction energies between inhibitor G39 (**11**) and NA residues and the bound water molecule in structure 2qwk. Product of the interaction energies (c, electrostatic; d, van der Waals) and the (real) PLS coefficients.

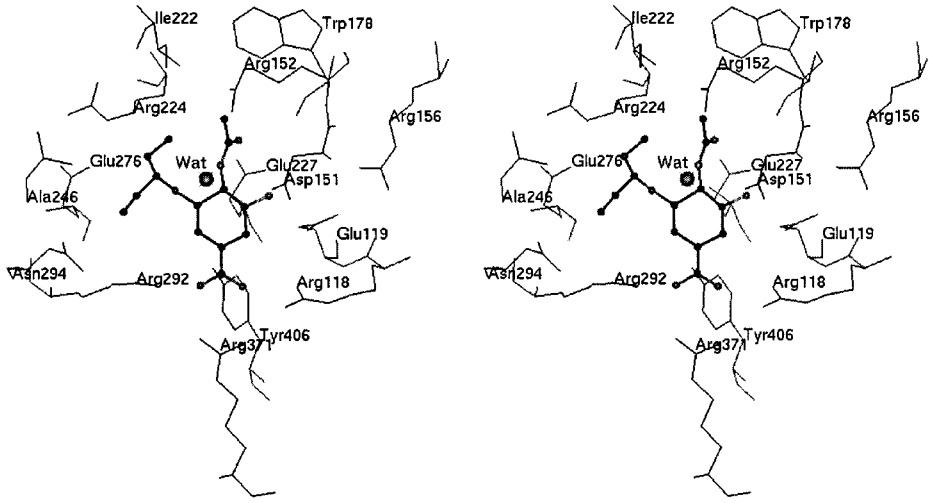


Figure 6. Stereoview of the active site of NA (structure 2qwk). The residues are labeled. The inhibitor G39 (**11**) is shown in ball-and-stick. The bound water molecule is shown by a ball.

final model (see Table 2). This N9 model exhibits much better fitting and predictive ability than the N2 + N9 model. At the optimal number of latent variables of 4, it gives an R^2 value of 0.94, SDEC of 0.45, Q^2 of 0.84, and SDEP of 0.75. The residues with PLS coefficients above the same thresholds as applied before are almost the same as those in the N2 + N9 model with the following two exceptions: Glu227 appears in the list of residues with a significant van der Waals contribution and Tyr406 disappears from the list of residues with a significant electrostatic contribution. It is difficult to explain these differences from protein structural differences between N2 NA and N9 NA as the orientations of Glu227 and Tyr406 are almost the same in N2 and N9. A possible explanation could be the differences between the inhibitors in the N2 + N9 data set and the N9 data set.

Extension of the Model. The three sialic acid (**1**) complexes (2qwb, 1mwe, 2bat) and the aPANA (**4**) complex (1inw) were not included in the above models as they are outliers. As shown in Table 1 and Figure 2, their activities are dramatically overpredicted by the N2

+ N9 model. Sialic acid (2bat) was also an outlier in the IC_{50} computation of Muegge.[27] Its IC_{50} value was overpredicted in his PMF scoring and so too was that of aPANA.

Several explanations can be provided for the overestimation of the COMBINE model. First, it is known that sialic acid exists in solution as a mixture of α- and β-anomers in a ratio of approximately 1:10,[46] and only the α-anomer binds in the active site. Considering this fact, the activity of α-sialic acid should be 10 times higher than the measured value for the mixture. Second, a conformational change of the ring is required for α-sialic acid binding. The pyranose ring has to convert from a chair form in solution to the less stable boat form upon binding (see Scheme 1). The binding energy would be reduced by this unfavorable conformational change, but the conformational change was not explicitly considered in the above model. PANA has no such enantiomorphism problem,[47] but the same ring conformational change is required for aPANA to bind to the NA active site.

We tried to incorporate the inhibitor intramolecular energy changes into the COMBINE analysis to account for the conformational changes of the ring. To model the unbound solution state of each inhibitor, a single conformation of each inhibitor was surrounded by explicit water molecules and energy-minimized. The intramolecular energy difference between an inhibitor in solution and in complex was used as an additional energy term to correlate with the activity by PLS, but sialic acid (**1**) and aPANA (**4**) remained outliers. A possible reason for this is that the entropic differences due to occupancy of additional conformations in solution were not taken into account in this simple procedure. We also attempted to incorporate estimates of desolvation energies based on the changes in solvent accessible surface area of the inhibitors upon binding, but this did not improve the model or alleviate the problem of the outliers. Another possibility, which was not investigated in the present study, would be to introduce electrostatic desolvation effects computed with a Poisson−Boltzmann model into the model.[30]

Design of New Inhibitors. On the basis of the above COMBINE analysis, some suggestions can be given for the design of new inhibitors. The triarginyl cluster of Arg118, Arg292, and Arg371 at the C1-pocket is the predominant factor for orienting and stabilizing inhibitors. A negatively charged group that makes strong charge−charge interactions with the triarginyl pocket is highly favorable for binding and is present in all the inhibitors studied. To replace the carboxylic and phosphonic groups in the current inhibitors, sulfinic (−SOOH) or sulfonic groups ($-SO_2OH$) are potential candidate substituents for the C1-pocket.

In contrast to the C1-pocket, a positively charged group is favored for binding in the C4-pocket by the interactions with Asp151 and Arg156. Accordingly, some of the current inhibitors that lack C4 positively charged groups could be modified to satisfy these interactions. For example, a higher activity is predicted if an amino group is substituted at the C4-position of bcx-140 (**18**).

In the C5-pocket, a hydrogen bond acceptor, e.g. a carbonyl group (−CO−), could strengthen binding by forming a hydrogen bond with Arg152. In addition, as

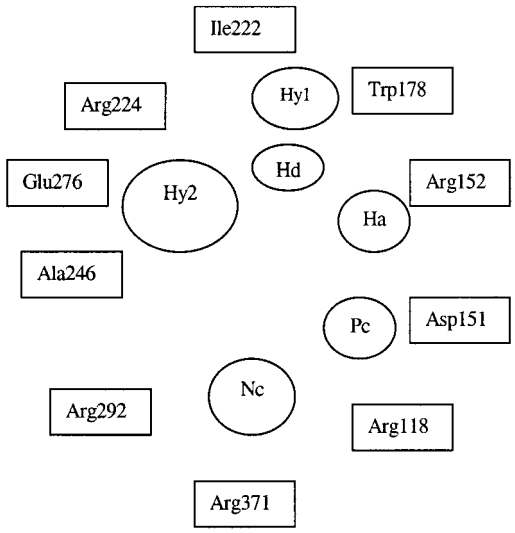


Figure 7. Diagram of important structural features (○) for a strong inhibitor and corresponding NA residues (□). Nc, negatively charged group; Pc, positively charged group; Ha, hydrogen bond acceptor; Hd, hydrogen bond donor; Hy1, small hydrophobic group; Hy2, large hydrophobic group.

the important bound water molecule acts as the bridge linking an inhibitor and NA residues by hydrogen bonds, a possible variant on this point is to replace or incorporate this water molecule into an inhibitor like one of the hydroxyl groups in QWL (**30**). There are two regions in which hydrophobic groups can be introduced to enhance binding. One is the small hydrophobic pocket formed by Trp178 and Ile222 as well as the side chain of Arg224. A larger one is created by rotating Glu276 as described above.

The important structural features for a strong inhibitor and corresponding protein residues are schematically shown in Figure 7. As the inhibitor rings have no direct interactions with NA and act only as scaffolds to orient the attached groups, novel inhibitors can be designed by constructing novel frame structures. For example, the new inhibitor bcx-1812 (**31**), which is currently in the clinical trials, has a five-membered ring framework. The experimental pIC_{50} value is in the range of 8.85−10 against 15 different strains of type A NA.[18] This inhibitor was published after we had completed our COMBINE model. To investigate how accurately our COMBINE model predicts the activity of this new inhibitor with a novel framework, we docked this inhibitor into the N9 protein of the complex 2qwk to create a N9 complex (rw9) and into the N2 protein of the complex 1ivf (with Glu276 of the N2 protein manually rotated to form a salt link to Arg224 before docking) to create a N2 complex (rw2). The modeling protocol was the same as described in the section Molecular Mechanics Modeling. The pIC_{50} values predicted by the COMBINE model were 8.36 for the N9 subtype and 7.13 for the N2 subtype. The N9 prediction is very close to the experimental value, while the underprediction for the N2 subtype is consistent with underpredictions for other inhibitors in the COMBINE model, e.g. for the inhibitor GNA (**7**) in the N2 complex dk1 (see Table 1).

The above discussion mainly refers to wild-type NA, but as NA is a target with a high mutation frequency, resistance to viral mutation should also be considered. For example, in this data set, the Arg292Lys mutation

resulted in a general reduction of inhibition activity due to reduced charge−charge interactions in the C1-pocket. On the other hand, the crystal structure[48] (2qwh) of the **11**:N9 mutant complex showed that the C6-hydrophobic pocket failed to form because the mutant residue Lys292 hindered the rotation of Glu276 and thus resulted in the largest resistance being shown to the inhibitor G39 (**11**). Therefore, to minimize viral resistance, inhibitor binding affinity should avoid strong dependence on Arg292. Another strategy to resist viral mutation is to design inhibitors targeting residues where mutations are highly deleterious to NA. According to the experimental work of Ghate and Air,[49] mutations at Tyr406, Glu276, and Asp151 have a drastic effect on NA activity: These residues would be good targets to design inhibitors with resistance to viral mutation.

Concluding Remarks

A series of NA−inhibitor complexes has been studied by COMBINE analysis. Despite the complexity of the system, predictive and robust QSAR models have been obtained by correlating the pIC_{50} values with the van der Waals and electrostatic interactions between the inhibitor and each protein residue and the bound water molecule. In the COMBINE models, the differences in inhibitory activity for the set of inhibitors are mainly determined by interactions with 12 active site residues and 1 bound water molecule. Accordingly, strong inhibitors should have structural features that make favorable interactions with these protein residues. They are a negatively charged group at the C1-pocket, a positively charged group at the C4-pocket, a hydrogen bond acceptor and a small hydrophobic group at the C5-pocket, a large hydrophobic group at the C6-pocket, and a hydrogen bond donor at the water position. The COMBINE model has been derived for N2 and N9 subtypes of influenza A and should be a useful aid for the optimization and de novo design of NA inhibitors.

Acknowledgment. We thank Dr. Y. S. Babu of BioCryst Pharmaceuticals, Inc. for providing the crystal structures of eight neuraminidase−inhibitor complexes. We thank Dr. G. Cruciani for providing the GOLPE program and Dr. A. Ortiz for providing the COMBINE program. T.W. is the recipient of a DAAD postdoctoral fellowship.

Supporting Information Available: All topology and coordinate files. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Pinto, L. H.; Holsinger, L. J.; Lamb, R. A. Influenza Virus M2 protein has Ion Channel Activity. *Cell* **1992**, *69*, 517−528.
(2) Colman, P. M.; Varghese, J. N.; Laver, W. G. Structure of the Catalytic and Antigenic Sites in Influenza Virus Neuraminidase. *Nature* **1983**, *303*, 41−44.
(3) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. Structure of the Haemagglutinin Membrane Glycoprotein of Influenza Virus at 3 Å Resolution. *Nature* **1981**, *289*, 366−373.
(4) Calfee, D. P.; Hayden, F. G. New Approaches to Influenza Chemotherapy: Neuraminidase Inhibitors. *Drugs* **1998**, *56*, 537−553.
(5) Kim, C. U.; Chen, X.; Mendel, D. B. Neuraminidase Inhibitors as Anti-influenza Virus Agents. *Antiviral Chem. Chemother.* **1999**, *10*, 141−154.
(6) Gubareva, L. V.; Kaiser, L.; Hayden, F. G. Influenza Virus Neuraminidase Inhibitors. *Lancet* **2000**, *355*, 827−825.
(7) Colman, P. M. Influenza Virus Neuraminidase: Structure, Antibodies, and Inhibitors. *Protein Sci.* **1994**, *3*, 1687−1696.
(8) Meindl, P.; Bodo, G.; Palese, P.; Schulman, J.; Tuppy, H. Inhibition of Neuraminidase Activity by Derivatives of 2-deoxy-2,3-dehydro-*N*-acetylneuraminic Acid. *Virology* **1974**, *58*, 457−463.
(9) Von Itzstein, M.; Wu, W. Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Van Phan, T.; Smythe, M.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational Design of Potent Sialidase-based Inhibitors of Influenza Virus Replication. *Nature* **1993**, *363*, 418−423.
(10) Kim, C. U.; Lew, W.; Williams, M. A.; Wu, H.; Zhang, L.; Chen, X.; Escarpe, P. A.; Mendel, D. B.; Laver, W. G.; Stevens, R. C. Structure−Activity Relationship Studies of Novel Carbocyclic Influenza Neuraminidase Inhibitors. *J. Med. Chem.* **1998**, *41*, 2451−2460.
(11) Taylor, N. R.; Cleasby, A.; Singh, O.; Skarynski, T.; Wonacott, A. J.; Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. Dihydropyrancarboxamides related to Zanamivir: A New Series of Inhibitors of Influenza Virus Sialidases. 2. Crystallographic and Molecular Modelling Study of Complexes of 4-Amino-4H-pyran-6-carboxamides and Sialidase from Influenza Virus Types A and B. *J. Med. Chem.* **1998**, *41*, 798−807.
(12) White, C. L.; Janakiraman, M. N.; Laver, W. G.; Philippon, C.; Vasella, A.; Air, G. M.; Luo, M. A Sialic Acid-derived Phosphonate Analogue Inhibits Different Strains of Influenza Virus Neuraminidase with Different Efficencies. *J. Mol. Biol.* **1995**, *245*, 623−634.
(13) Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Ali, S. M.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E. A.; Finley, J.; Luo, M.; Air, G. M.; Laver, G. W. Potent Inhibition of Influenza Sialidase by a Benzoic Acid Containing a 2-Pyrrolidinone Substituent. *J. Med. Chem.* **1999**, *42*, 2332−2343.
(14) Singh, S.; Jedrzejas, M. J.; Air, G. M.; Luo, M.; Laver, W. G.; Brouillette, W. J. Structure-based Inhibitors of Influenza Virus Sialidase. A Benzoic Acid with Novel Interaction. *J. Med. Chem.* **1995**, *38*, 3217−3225.
(15) Chand, P.; Babu, Y. S.; Bantia, S.; Chu, N.; Cole, B.; Kotian, P. L.; Laver, W. G.; Montagomery, J. A.; Pathak, V. P.; Petty, S. L.; Shrout, D. P.; Walsh, D. A.; Walsh, G. M. Design and Synthesis of Benzoic Acid Derivatives as Inflluenza Neuraminidase Inhibitors Using Structure-based Drug Design. *J. Med. Chem.* **1997**, *40*, 4030−4052.
(16) Sudbeck, E. A.; Jedrzejas, M. L.; Singh, S.; Brouillette, W. J.; Air, G. M.; Laver, W. G.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Luo, M. Guanidinobenzoic Acid Inhibitors of Influenza Virus Neuraminidase. *J. Mol. Biol.* **1997**, *267*, 587−594.
(17) Brouillette, W. J.; Atigadda, V. R.; Luo, M.; Air, G. M.; Babu, Y. S.; Bantia, S. Design of Benzoic Acid Inhibitors of Influenza Neuraminidase Containing A Cyclic Substitution for the *N*-acetyl Grouping. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1901−1906.
(18) Babu, Y. S.; Chand, P.; Bantia, S.; Montgonery, J. A. BCX-1812-(RWJ-270201): Discovery of a Novel, Highly Potent, Orally Active, and Selective Influenza Neuraminidase Inhibitor Through Structure-based Drug Design. *J. Med. Chem.* **2000**, *43*, 3482−3486.
(19) Wade, R. C. 'Flu' and structure-based Drug Design. *Structure* **1997**, *5*, 1139−1145.
(20) von Itzstein, M.; Dyason, J. C.; Oliver, S. W.; White, H. F.; Wu, W. Y.; Kok, G. B.; Pegg, M. S. A Study of the Active Site of Influenza Virus Sialidase: An Approach to the Rational Design of Novel Anti-influenza Drugs. *J. Med. Chem.* **1996**, *39*, 388−391.
(21) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857
(22) Taylor, N. R.; von Itzstein, M. A Structural and Energetics Analysis of the Binding of a Series of *N*-acetylneuraminic-acid-based Inhibitors to Influenza Virus Sialidase. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 233−246.
(23) Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Air, G. M.; Luo, M. A Strategy for Theoretical Binding Constant, Ki, Calculations for Neuraminidase Aromatic Inhibitors Designed on the Basis of the Active Site Structure of Influenza Virus Neuraminidase. *Proteins: Struct. Funct. Genet.* **1995**, *23*, 264−277.
(24) Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method. *J. Med. Chem.* **1999**, *42*, 5142−5152.
(25) Aqvist, J.; Medina, C.; Samuelsson, J. E. New Method for Predicting Binding Affinity in Computer-aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.
(26) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem.* **1999**, *42*, 791−804.

(27) Muegge, I. The Effect of Small Changes in Protein Structure on Predicted Binding Modes of Known Inhibitors of Influenza Virus Neuraminidase: PMF-scoring in DOCK4. *Med. Chem. Res.* **1999**, *9*, 490−495.
(28) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.* **1995**, *38*, 2681−2691.
(29) Ortiz, A. R.; Pastor, M.; Palomer, A.; Cruciani, G.; Gago, F.; Wade, R. C. Reliability of Comparative Molecular Field Analysis Models: Effects of Data Scaling and Variable Selection Using a Set of Human Synovial Fluid Phospholipase A2 Inhibitors. *J. Med. Chem.* **1997**, *40*, 1136−1148
(30) Perez, C.; Pastor, M.; Ortiz, A. R.; Gago, F. Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-based Drug Design. *J. Med. Chem.* **1998**, *41*, 836−852.
(31) Pastor, M.; Gago, F.; Cruciani, G. Comparative Binding Energy (COMBINE) Analysis on a series of Glycogen Phosphorylase Inhibitors. Comparison with GRID/GOLPE Models. In *Molecular Modeling and Prediction of Bioactivity*; Gundertofte, K., Jorgensen, F. S., Eds.; Kluwer: New York, 2000; pp 329−330.
(32) Lozano, J. J.; Pastor, M.; Cruciani, G.; Gaedt, K.; Centeno, N. B.; Gago, F.; Sanz, F. 3D-QSAR Methods on the Basis of Ligand Receptor Complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 Ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 341−353.
(33) Tomic, S.; Nilsson, L.; Wade, R. C. Nuclear Receptor-DNA Binding Specificity: A COMBINE and Free-Wilson QSAR Analysis. *J. Med. Chem.* **2000**, *43*, 1780−1792.
(34) AUTODOCK: Automated Docking of Flexible Ligands to Receptors, version 3.0; The Scripps Research Institute, La Jolla, CA, 1999.
(35) Mckimm-Breschkin, J. L.; Sahasrabudhe, A.; Blick, T. J.; Mcdonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a Conserved Residue in the Influenza Virus Neuraminidase Active Site Decrease Sensitivity to Neu5Ac2en-derived Inhibitors. *J. Virol.* **1998**, *72*, 2456−2462.
(36) Cheng, Y.; Prusoff, W. H. The relationship between Inhibition constant K*i* and the Concentratation of Inhibitor Which Causes 50% Inhibition I_{50} of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(37) Vriend, G. WHATIF: A Molecular Modelling and Drug Design Program. *J. Mol. Graph.* **1990**, *8*, 52−56.
(38) Hooft, R. W. W.; Sander, C.; Vriend, G. Positioning Hydrogen Atoms by Optimizing Hydrogen Bond Networks in Protein Structures. *Proteins* **1996**, *26*, 363−376. Pinto, L. H.; Holsinger, L. J.; Lamb, R. A. Influenza Virus M2 protein has Ion Channel Activity. *Cell* **1992**, *69*, 517−528.
(39) AMBER: Assisted Model Building with Energy Refinement, version 5.0; Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, 1997.
(40) InsightII; Molecular Simulations Inc., San Diego, CA, 1997.
(41) Cornell, W. D.; Cieplak, P.; Payly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217. Modified version in 1997.
(43) SYBYL package, version 6.5; Tripos Associates, Inc., St. Louis, MO, 1998.
(44) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9−20.
(45) Blick, T. J.; Tiong, T.; Sahasrabudhe, A.; Varghese, J. N.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Mckimm-Breschkin, J. L. Generation and Characterization of an Influenza Virus Neuraminidase Variant with Decreased Sensitivity to the Neuraminidase-Specific Inhibitor 4-Guanidino-Neu5Ac2en. *Virology* **1995**, *214*, 475−484.
(46) Jaques, L. W.; Brown, E. B.; Barrett, J. M.; Brey, W. S., Jr.; Weltner, W., Jr. Sialic Acid: a Calcium-binding Carbohydrate. *J. Biol. Chem.* **1977**, *252*, 4533−4538.
(47) Wallimann, K.; Vasella, A. Phosphonic-Acid Analogues of the N-Acetyl-2-deoxyneuraminic Acids: Synthesis and Inhibition of Vibrio Cholerae Sialidase. *Helv. Chim. Acta* **1990**, *73*, 1359−1372.
(48) Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; Mckimm-Breschkin, J. L.; Colman, P. M. Drug Design Against a Shifting Target: a Structural Basis for Resistance to Inhibitors in a Variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735−746.
(49) Ghate, A. A.; Air, G. M. Site-directed Mutagenesis of Catalytic Residues of Influenza Virus Neuraminidase as an Aid to Drug Design. *Eur. J. Biochem.* **1998**, *258*, 320−331.

JM001070J